ENCODING IN THE DARK GRAND CHALLENGE: AN OVERVIEW

Nantheera Anantrasirichai, Fan Zhang, Alexandra Malyugina, Paul Hill, and Angeliki Katsenou

Visual Information Laboratory, University of Bristol, UK

ABSTRACT

A big part of the video content we consume from video providers consists of genres featuring low-light aesthetics. Low light sequences have special characteristics, such as spatio-temporal varying acquisition noise and light flickering, that make the encoding process challenging. To deal with the spatio-temporal incoherent noise, higher bitrates are used to achieve high objective quality. Additionally, the quality assessment metrics and methods have not been designed, trained or tested for this type of content. This has inspired us to trigger research in that area and propose a Grand Challenge on encoding low-light video sequences. In this paper, we present an overview of the proposed challenge, and test state-of-the-art methods that will be part of the benchmark methods at the stage of the participants' deliverable assessment. From this exploration, our results show that VVC already achieves a high performance compared to simply denoising the video source prior to encoding. Moreover, the quality of the video streams can be further improved by employing a post-processing image enhancement method.

Index Terms— Video coding, VVC, denoising, quality metrics, low light scenes.

1. INTRODUCTION

Last year, the HBO Game of Thrones episode, entitled "The Long Night", received a lot of controversial reviews because it was shot in low light and many fans complained about the picture quality [1]. Low light scenes often come with acquisition noise, which not only disturbs the viewers, but also brings special characteristics to video compression. This noise appears randomly in time, possibly creating visibly temporal flickering. These type of videos are often encountered in cinema as a result of artistic perspective or the nature of a scene. Other examples include shots of wildlife (e.g. mobula rays at night in Blue Planet II), concerts, shows, surveillance camera footage and more. In this context, we study the encoding of videos captured in low-light using state-of-the-art methods that has inspired us to organise a video coding Grand Challenge within IEEE ICME2020.

Noise can be introduced during video acquisition, recording and processing. Not only visually unpleasant, noise also affects the performance of intra and inter prediction in video compression, causing the encoder to inefficiently spend bits to represent this noise, especially at low compression levels. Currently, noise reduction techniques are usually employed for film-grain noise in the creative industry during a pre-encoding phase with the aim of improving compression performance. Later, synthetic noise is superimposed at the decoded video sequence [2, 3]. However, film-grain noise is a special case as it is considered part of the artistic effect that enhances the natural appearance of the video and the viewers are quite comfortable with it. On the other hand, there are other types of noise that the viewers do not like and perceive in many cases as 'low' quality. In this paper, we consider types of noise that are undesirable for viewers, consequently, we do not attempt to restore the noise after the decoding. The simplest technique of noise reduction is a weighted averaging technique performed in a temporal sliding window—a.k.a. moving average filter [4]. More sophisticated methods include adaptive spatio-temporal smoothing through anisotropic filtering [5], nonlocal transform-domain group filtering [6], Kalman-bilateral mixture model [7], and spatio-temporal patch-based filtering [8]. Recent work employs the popular deep learning approach. For example, a residual noise map is estimated in the Denoising Convolutional Neural Network (DnCNN) method [9] for image based denoising, and for video based denoising, a spatial and temporal network are concatenated where the latter handles brightness changes and temporal inconsistencies [10]. VNLnet combines a non-local patch search module with DnCNN. The first part extracts features, and the latter mitigates the remaining noise [11].

Another direction in video coding is to perform denoising in the loop, such as [12, 13]. In-loop filters have been proposed and adopted by recent video coding standards (e.g. [14]). The most popular examples of in-loop filters are the adaptive deblocking, the adaptive loop, the sample adaptive offset, and Convolutional Neural Network (CNN) based filters. These filters however are meaningful in light compression, where they contribute towards better intra prediction. In heavy compression, quantization filters out the noise [12]. CNN-based methods are also popular for postprocessing and can provide significant image enhancement, leading to better final rate-distortion performance with significantly lower complexity [15, 16].

In this paper, we briefly describe the Grand Challenge we are organising on encoding low-light sequences and we present some solutions using state-of-the-art methods to improve the compression performance. As quality metrics play an important role in the assessment of both anchors and the challenge participant submissions, we explore the performance of state-of-the-art objective Image and Video Quality Assessment (IQA/VQA) metrics on low-light video content. We first test the performance of recent video codec standard—Versatile Video Coding (VVC) on this type of content. Subsequently, we implement and present two possible workflows combining the codec and the denoising module: i) applying denoising before coding (pre-processing methods), and ii) applying image enhancement after coding (post-processing methods).

The remainder of this paper is organised as follows. Section 2 discusses the difficulties of encoding dark-scene videos. Then the evaluation metrics and the benchmark methods are described in Section 3 and Section 4, respectively. Section 5 presents the evaluation results with discussion. Finally, a conclusion is outlined in Section 6.

2. LOW LIGHT VIDEOS AND THE DIFFICULTY IN ENCODING

Inspired by all above, we are organising a challenge on encoding low-light captured videos within IEEE ICME 2020. This challenge intends to identify technology that improves the perceptual quality of compressed low-light videos beyond the current state-of-the-art

978-1-7281-1485-9/20/$31.00

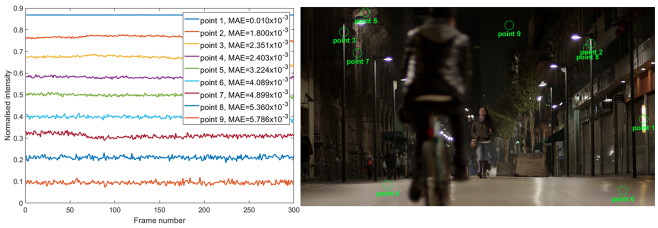


Fig. 1. Normalised intensity values (left) at various background locations (right) in each time frame of sequence S3.

performance of the most recent coding standards, such as HEVC, AV1, VVC, etc. The challenge encourages the exploration of novel technologies applied within existing video codecs (e.g. in loop filters), or/and prior to encoding processing methods (e.g. denoising) and/or post-processing methods (e.g. de-blocking).

The darker the light conditions, the harder it is to effectively capture video. However, such conditions can be very aesthetically and artistically rewarding. Moreover, shooting with low light also means that ISO noise will be present and it will be temporally incoherent. Frequently, light sources are not always consistent throughout the shot, posing additional challenges. To illustrate an example of the temporal variation of intensity values in a dark sequence, we plotted the Y component as a time series for different intensity values in Fig. 1. We computed the mean absolute errors (MAEs) from the smooth curves (an average of a sliding window with the size of 20 frames), and found that the darker areas have larger MAEs, implying higher noise levels. The brighter pixels have lower variance, but there is still a small temporal variation despite the fact that all points are part of a static background. This is because the dark scenes generally contain several different light sources. Although this temporal variation may be unnoticeable to the viewers, it could significantly deteriorate the encoding performance and make a high target quality very expensive in terms of bitrate.

2.1. Dataset

For the Grand Challenge dataset[1], we have selected 6 Full High Definition (FHD) (1920×1080) YUV sequences with 10 bits depth and 420 colour sampling. All of the sequences except for Campfire (30fps) are at 60fps. The detailed list and sources of the test dataset is reported in Fig 2. We have carefully selected the sequences of different content, namely to include static or moving scenes, different types of motion, areas of interest, different luminance distribution including one or more light sources, varying background, and also dynamic textures (S5-S6). Sequences S1-S5 were captured by professionals, all are free for research purposes and all have been used in video coding standardisation activities or recent publications. Sequence S6 was captured by University of Bristol for a study of dynamic textures.

2.2. Video Coding

Since its launch in 2004, H.264/AVC [21] has been the most deployed video coding standard, despite the fact that its successor, H.265/HEVC [22, 23, 24] released in 2013, as it provides enhanced coding performance. The MPEG standardisation body is currently working towards the next generation video coding standard, Versatile Video Coding (VVC). It has been named versatile, as it is supporting immersive formats (higher spatial resolutions, higher dynamic range and 360° videos). Recently, there has been increased

[1]https://ieee-dataport.org/competitions/encoding-dark

Fig. 2. Sample frames from the test dataset.

activity in the video coding technology industry with the aim to develop open-source royalty-free video codecs, particularly by the Alliance for Open Media (AOMedia). In 2018, AV1 (AOMedia Video 1) [25] was released as a competitor to HEVC. AV1 was primarily based on Google's video codec VP9 [26] and has comparable performance to HEVC [27, 28, 29, 30].

3. EVALUATION METRICS

In order to perform quality assessment in video coding, two standard methodologies are usually employed: the computation of objective IQA/VQA metrics and subjective quality assessment. In this paper, we only perform objective quality assessment and leave the subjective evaluation, which is generally time-intensive, to be performed after we receive all participants' deliverables.[2] Particularly, we explore the literature and evaluate commonly employed full reference and no reference IQA/VQA metrics. We decided to test both types of metrics due to the fact that: firstly, in video coding usually full-reference IQA/VQA metrics are used since the original video source is available (in most cases) and, secondly, in video denoising applications no reference metrics are typically used as there is no 'reference' clean (denoised) sequence available.

3.1. Full Reference Metrics

Full Reference (FR) IQA/VQA metrics have been traditionally employed for video compression purposes and consist of a wide variety including Peak Signal to Noise Ratio (PSNR), that takes into account the Contrast Sensitivity Function (CSF) and the between-coefficient contrast masking of Discrete Cosine Transform (DCT) basis functions (PSNR-HVSM) [31], Structural Similarity Index (SSIM) [32] and multi-scale SSIM (MS-SSIM) [33], Visual Information Fidelity measure (VIF) [34] that is often employed as a feature, Video Quality Metric (VQM) [35], Spatio-Temporal Most Apparent Distortion model (ST-MAD) [36] and Video Multi-method Assessment Fusion (VMAF) metric (using model vmaf_v0.6.1.pkl, which was trained for FHD content) [37]. PSNR, PSNR-HVSM, SSIM, MS-SSIM, and VIF are commonly used IQA metrics, while VQM, ST-MAD and VMAF are VQA methods. All of these metrics have their strengths and weaknesses in terms of their correlation to subjective scores and complexity. None of these metrics have been rigorously tested on low-light image or video compressed content.

3.2. No Reference Metrics

No Reference (NR) IQA/VQA metrics are usually employed when the reference source sequences are not available (e.g. in user-generated content) or when the capture artefacts are dominant and

[2]An extensive report of both objective and subjective quality assessment, as well as their correlations to the subjective scores will be presented during the special session in the conference.

Table 1. Test sequences and target bit rate points (in kbps).

No.	Sequences	Target bit rates [kbps]			
		Rate1	Rate2	Rate3	Rate4
S1	ElFuente-Palacio	100	170	300	500
S2	ElFuente-Cars	85	150	280	540
S3	ElFuente-Cyclist	70	120	210	400
S4	Chimera-Dinner	50	70	100	200
S5	Campfire	640	1300	2500	4500
S6	SmokeClear	220	400	700	1400

the reference sequence is considered impaired. Several different no reference metrics have been proposed in the literature. The JPEG [38] and JPEG2000 (JP2K) [39] quality scores were two of the first no reference IQA methods introduced. The first attempts to align image quality with HVS (Human Vision System) perception by characterising blockiness and blurring. Since then a variety of different no reference quality metrics have been proposed in the literature, such as the Anisotropic Quality Index (AQI) [40], the Blind Image Quality Index (BIQI) [41], the Contrast Enhancement (CEIQ) employed to measure contrast distortion in [42], the Naturaleness Image Quality Evaluator (NIQE) [43], the (PIQE) [44], the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [45], the Video BLIINDS [46], and the Two-Level Video Quality Metric (TLVQM) [47]. Most of the aforementioned metrics have been designed based on Natural Scene Statistics theory, taking into account features such as contrast, intensity, colour, spatial and temporal correlation of frequencies and their statistical distributions. Lately, a lot of learning-based methods have been proposed, such as BRISQUE, TLVQM and more. However, none of these methods have been trained on low-light content.

4. EXPERIMENT CONFIGURATIONS

4.1. Video Coding Conditions and Anchors

We followed one of the coding constraint cases in JVET Joint Call for Proposals on Video Compression with Capability beyond HEVC [48]:

- No more than 16 frames of structural delay, e.g. "group of pictures" (GOP) of 16.
- A random access interval of 1.1 seconds or less - defined as 32 pictures or less for a video sequence with a frame rate of 30 frames per second, 64 pictures or less for a video sequence with a frame rate of 64 frames per second.

We have selected a set of bitrates that are different for each sequence as shown in Table 1. These target bitrates were selected after a small scale expert study, so that the perceived quality of the encoded anchor bitstreams will uniformly cover the quality scale from "Poor" up to "Good" and give to the participants room for improvement in the scale of perceived quality.
Anchor bitstreams are generated using the VVC VTM 7.0 software, which uses static Quantisation Parameter (QP) configurations. A one-time change of QP, through the encoder parameter of `QPIncrementFrame (-qpif)` in VTM, was used to achieve the defined target bit rates for some bitstreams. The bitrates of produced anchor bitstreams do not exceed the target rate points by 3%. We did not use rate control to hit the target bitrate as this compromises the quality of the resulting bitstream.

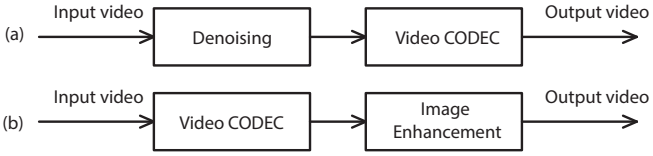


Fig. 3. Diagrams of two processes for Encoding in the Dark: (a) preprocessing and (b) postprocessing methods.

4.2. Benchmark methods

Two workflows examined and compared are shown in Fig 3. The preprocessing method applies denoising at the encoder, whilst the postprocessing method applied image enhancement at the decoder. Both of them operate on frame-by-frame basis.

4.2.1. Preprocessing method with Denoising

Learning-based denoising methods have been proven to outperform traditional filtering techniques in both quality and speed. Amongst these, Denoising CNN (DnCNN) [9] has become popular due to its reconstruction performance, simple implementation and computational speed [49]. The utilised DnCNN architecture comprises 17 convolutional layers combined with batch normalisation and a ReLU activation [9]. The network does not include pooling layers and therefore mainly extracts low-level features, fundamental for modelling noise in the image. In this paper, we used an adapted model trained on colour images with synthetic Gaussian noise.

We investigated the levels of temporal noise in denoised videos similar to that examined in Section 2 shown in Fig. 1. The MAEs ($\times 10^{-3}$) of the nine points in the denoised sequence are 0.012, 1.678, 2.147, 2.167, 3.012, 3.869, 4.552, 4.697, and 4.886, respectively. Eight of nine are less than the values of the original noisy videos, particularly in dark areas. This implies that temporal variation is reduced after denoising, which results in more precise motion estimation with smaller residuals in the coding process.

4.2.2. Postprocessing Method with Image Enhancement

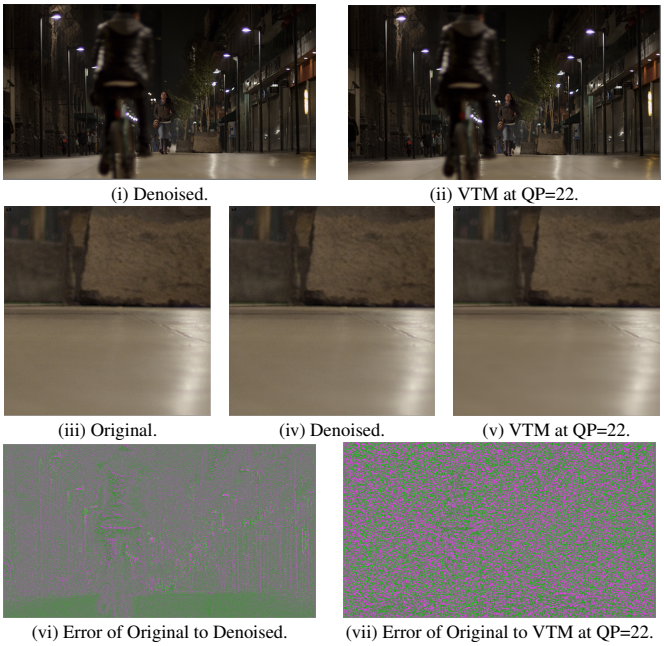
Postprocessing is commonly applied at the video decoder, on the reconstructed frames, to reduce various coding artefacts and enhance visual quality. Here, we employed the CNN-based postprocessing method proposed by Zhang et al. [16], which has been reported to offer significant coding gains over VVC. Its network architecture was modified based on the generator (SRResNet) of SRGAN [50]. It contains $2N$+2 convolutional layers, all of which have 3×3 kernels, 64 feature maps and a stride value of 1, except the last convolutional layer (with 3 feature maps instead). Between the first and the last convolutional layers, there are N identical residual blocks, each of which contains two convolutional layers and a parametric ReLU activation function in between them. Additional skip connections are employed (i) between the input of the first residual block and the output of the N^{th} residual block (ii) between the input of the CNN and the output of the last convolutional layer. Here we used N=16.

5. RESULTS AND DISCUSSION

In the next subsections, we first discuss the results of the tested methods and then explain the limitations and challenges that will be part of our future work.

5.1. Results

In the main phase of the evaluation, we executed the methods described in Section 4. The experiments were performed on a clus-

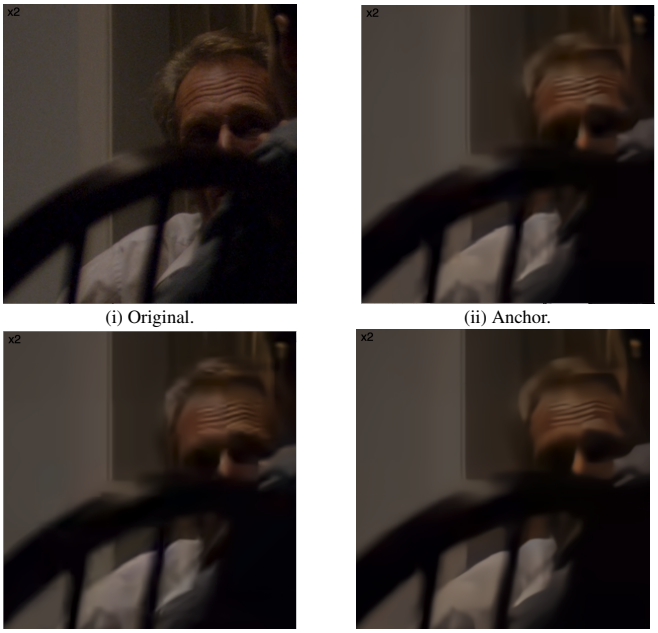


(i) Denoised. (ii) VTM at QP=22.

(iii) Original. (iv) Denoised. (v) VTM at QP=22.

(vi) Error of Original to Denoised. (vii) Error of Original to VTM at QP=22.

Fig. 4. Example of visual inspection of S3 ElFuente-Cyclist sequence of the denoising effect by applying denoising or by just encoding with VVC.



(i) Original. (ii) Anchor.

(iii) Preprocessed. (iv) Postprocessed.

Fig. 5. Example of visual inspection of S4 ElFuente-Dinner sequence comparing the tested methods at Rate 4.

ter computer, in which each node contains 2.4 GHz Broadwell Intel CPUs, 128GB RAM and 2x NVIDIA Tesla P100.

First, we perform a visual inspection of the denoising effect that a deep-learning based method can provide against the denoising result of video encoding at light compression (at low QPs). In Fig. 4 (i)-(ii), we illustrate an example of the first frame of the S3 ElFuente-Cyclist sequence after applying denoising, as described in Section 4.2.1, and the same frame after encoding the sequence with VTM7.0. Because the differences are almost imperceptible, we zoomed in at the same block and provide it at three different versions: in (iii) cropped from the original frame, in (iv) cropped from the denoised frame and in (v) cropped from the VTM encoded sequence. Figs. 4 (iii)-(iv) look very similar, while in (v) the compression effect is slightly apparent as the tiles look more uniform. The differences of the denoised and compressed frames from the respective original are confirmed by the visualisation of the frame difference in Figs. 4 (vi)-(vii). In these figures, the differences are amplified by a factor of 10. Green and pink colours represent pixel difference either positive or negative, and grey represents no difference. As can be seen, the differences are more intense for the compressed frame than the denoised.

From the resulting final YUV video sequences from the tested methods, we first plotted the rate-quality curves and then we computed the Bjøntegaard delta rate (BD-Rate) [51]. BD-Rate is widely used to calculate the coding efficiency between different coding technologies. The results of the pre- and post-processing methods are reported in Table 2, where negative and positive values represent gain and loss of the coding performance, respectively. We computed the BD-Rate values for the different rate-quality curves, by taking into account the IQA/VQA metrics that indicated the best performance in terms of monotonicity. Particularly, we employed two FR metrics, PSNR of Y component (PSNR-Y) and VMAF, and two NR metrics, i.e. AQI and PIQE. PSNR is the most commonly used assessment method for video coding, while VMAF has been reported to offer better correlation with subjective opinions comparing to most of existing FR quality metrics [52]. Among most existing NF quality metrics, based on our preliminary study, the AQI and PIQE offer the best monotonicity characteristic against QP indices when coding the low-light sequences. Therefore these two metrics are employed here.

According to the PSNR-Y BD-Rate figures shown in Table 2, the postprocessing method appears to improve the VTM performance with an average gain of 1.8%. Conversely, applying denoising prior to encoding does not bring any gains but rather losses, with an average of 7.2% in terms of PSNR. Inspecting the losses per sequence, the higher PSNR-Y BD-Rate losses are reported for sequences S4 and S5, which are the sequences with lower contrast and areas of very low luminosity. A similar conclusion can be drawn from the VMAF BD-Rate figures with slightly higher average gains/losses.

Observing the NR BD-Rate figures in Table 2, we notice that AQI shows a similar trend to the FR BD-Rate figures, but with amplified gains/losses. On the contrary, the PIQE BD-Rate figures are quite different, showing large differences in rate-quality curves of the tested methods against the anchors.

The above observations emphasize the need for a subjective quality assessment, so that we can confirm that the metrics utilised are suitable for dark scenes content. To further support this, we are illustrating in Fig. 5 an example of cropped patches from sequence S4. As can be seen the quality is similar. The contrast in the preprocessed case is a bit higher and more details are preserved, for example the wrinkles in the forehead and the hair. In the case of the anchor and the postprocessed the forehead and the hair look flatter. All of these details though might not be noticeable during the video playout as they might be masked by motion.

5.2. Limitations and Open Issues

The results of the tested methods indicate some of the limitations and issues we encountered when we were preparing this Grand Challenge. The most important points can be summarised below:

Table 2. BD-Rate savings achieved by two benchmark methods over VVC VTM 7.0, assessed by four different quality metrics [51].

Metric	Method	S1	S2	S3	S4	S5	S6	Avg
$PSNR_Y$	Pre	1.1%	-0.5%	1.9%	8.0%	32.4%	0.33%	7.2%
	Post	-3.0 %	-1.4%	-1.5%	-1.1%	-2.9%	-1.1%	-1.8%
VMAF	Pre	3.5%	0.0%	6.3%	8.6%	14.0%	-1.9%	5.1%
	Post	-5.7%	-5.9%	-5.4%	-8.1%	-4.6%	1.5%	-4.7%
AQI	Pre	8.4%	-4.2%	-8.5%	10.9%	-39.3%	-3.6%	-6.0%
	Post	-9.5%	-11.9%	-14.9%	-20.5%	-10.0%	2.6%	-10.7%
PIQE	Pre	77.0%	118.3%	61.5%	53.9%	31.5%	-52.3%	48.3%
	Post	-54.2%	-62.2%	-58.2%	-21.5%	79.2%	-46.8%	-27.3%

- The lack of a large dataset with low-light video content certainly poses limitations on the effectiveness of learning-based methods. It is important to note that the models used in both pre- and post-processing frameworks are trained on generic datasets with natural scenes. We expect that their performance would be improved, were they retrained using dark scenes.
- The absence of a dataset with subjective evaluations poses another limitation as the considered FR/NR IQA/VQA metrics can not be fully validated on their performance on this type of content. First of all, without subjective evaluation data, we cannot answer the critical question of which type of metric is more suitable for low-light sequences. On one hand, a denoised problem is considered as referenceless, while on the other hand in the video coding pipeline you always have a source/reference sequence.
- The previous point also leads to the question of whether the metrics examined here are the best performing ones and whether the BD-Rate gains/losses reported in this paper are representative. We anticipate that with the extensive subjective evaluation, to be conducted within this Challenge, we will re-evaluate all metrics and conclude on the most suitable for this type of content.

6. CONCLUSION AND FUTURE WORK

This paper presents a study of encoding of low-light captured videos using contemporary methods, serving as the benchmark in the Grand Challenge within IEEE ICME2020. We carefully selected six dark-scene videos and provided anchors based on VVC. We then investigated available FR and NR evaluation metrics. Two workflows for encoding in the dark were examined: preprocessing with DnCNN-based denoising and postprocessing with a CNN-based image enhancement method. Experimental results show that the postprocessing framework outperforms the VTM alone by up to 1.8% on average based on PSNR-Y, whilst the state-of-the-art VTM encoder appears to not improve its performance (7.2%) by applying noise removal in the preprocessing framework.

In the future, the limitations and challenges identified will be further studied. Additional state-of-the-art methods will be benchmarked and compared against the Grand Challenge participants' deliverables. Furthermore, extensive experiments of subjective quality assessment will be performed to assess the perceptual gains of these methods. Another important outcome of this challenge is expected to be the evaluation of the correlation of the IQA/VQA methods with subjective quality for low-light sequences.

7. ACKNOWLEDGMENT

The work was supported in part by the Bristol & Bath R&D cluster (AH/S002936/1) to N. Anantrasirichai, in part by the Leverhulme Early Career Fellowship (ECF-2017-413) awarded to A. Katsenou. We also gratefully acknowledge the support of Facebook and Netflix for sponsoring the awards of the finalists.

8. REFERENCES

[1] A. Stout, "Game of thrones: Was the long night too dark?," *IBC*, May 2019.

[2] O. Stankiewicz, "Video coding technique with a parametric modelling of noise," *Opto-Electronics Review*, vol. 27, no. 3, pp. 241 – 251, 2019.

[3] B. T. Oh, S. Lei, and C.-J. Kuo, "Advanced film grain noise extraction and synthesis for high-definition video coding," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 19, no. 12, pp. 1717–1729, Dec 2009.

[4] A. A. Yahya, J. Tan, B. Su, and K. Liu, "Video denoising based on spatial-temporal filtering," in *6th Intern. Conf. on Digital Home*, Dec 2016, pp. 34–37.

[5] H. Malm, M. Oskarsson, E. Warrant, P. Clarberg, J. Hasselgren, and C. Lejdfors, "Adaptive enhancement and noise reduction in very low light-level video," in *IEEE ICCV*, Oct 2007, pp. 1–8.

[6] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, 2012.

[7] Zuo C, Liu Y, Tan X, Wang W, and Zhang M., "Video denoising based on a spatiotemporal kalman-bilateral mixture model," *ScientificWorldJournal*, vol. 438147, 2013.

[8] A. Buades and J. Duran, "Cfa video denoising and demosaicking chain via spatio-temporal patch-based filtering," *IEEE Trans. Circ. Syst. Video Tech.*, pp. 1–1, 2019.

[9] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Trans. on Image Processing*, vol. 26, no. 7, pp. 3142–3155, July 2017.

[10] M. Claus and J. van Gemert, "Videnn: Deep blind video denoising," in *CVPR workshop*, 2019.

[11] A. Davy, T. Ehret, J. Morel, P. Arias, and G. Facciolo, "A non-local cnn for video denoising," in *IEEE ICIP*, Sep. 2019, pp. 2409–2413.

[12] E. Wige, G. Yammine, P. Amon, A. Hutter, and A. Kaup, "In-loop noise-filtered prediction for high efficiency video coding," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 24, no. 7, pp. 1142–1155, 2014.

[13] M. Tang, Y. Han, J. Wen, and S. Yang, "HEVC-based motion compensated joint temporal-spatial video denoising," in *IEEE ICASSP*, March 2017, pp. 1797–1801.

[14] S. Wan, M.-Z. Wang, H. Gong, C.-Y. Zou, Y.-Z. Ma, J.-Y. Huo, Y.-F. Yu, and Y. Liu, "Ce10: Integrated in- loop filter based on cnn (tests 2.1, 2.2 and 2.3)," Tech. Rep. JVET meeting, no. JVET-O0079. ITU-T, ISO/IEC, ITU-R, 2019.

[15] M. Wang, S. Wan, H. Gong, and M. Ma, "Attention-Based Dual-Scale CNN In-Loop Filter for Versatile Video Coding," *IEEE Access*, vol. 7, pp. 145214–145226, 2019.

[16] F. Zhang, F. Chen, and D. R. Bull, "Enhancing VVC through CNN-based Post-Processing," in *IEEE ICME*, 2020.

[17] I Katsavounidis, "NETFLIX - "El Fuente" video sequence details and scenes," July 2015.

[18] I Katsavounidis, "NETFLIX - "Chimera" video sequence details and scenes," November 2015.

[19] Y. Zhu, L. Song, R. Xie, and W. Zhang, "SJTU 4K video subjective quality dataset for content adaptive bit rate estimation without encoding," in *Broadband Multimedia Systems and Broadcasting (BMSB), 2016 IEEE Intern. Symposium on*. IEEE, 2016, pp. 1–4.

[20] M. A. Papadopoulos, F. Zhang, D. Agrafiotis, and D. Bull, "A video texture database for perceptual compression and quality assessment," in *IEEE ICIP*, 2015, pp. 2781–2785.

[21] ITU-T Rec. H.264, "Advanced video coding for generic audiovisual services," 2005.

[22] ITU-T Rec. H.265, "High efficiency video coding," 2015.

[23] M. Wien, *High efficiency video coding*, Springer, 2015.

[24] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standard - including High Efficiency Video Coding (HEVC)," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 22, no. 12, pp. 1669–1684, 2012.

[25] AOM, "AOMedia Video 1 (AV1)," `https://github.com/AOMediaCodec`, 2019.

[26] "VP9 Video Codec," `https://www.webmproject.org/vp9/`.

[27] A. V. Katsenou, F. Zhang, M. Afonso, and D. R. Bull, "A subjective comparison of AV1 and HEVC for adaptive video streaming," in *IEEE ICIP*, 2019.

[28] P. Topiwala, M. Krishnan, and W. Dai, "Performance comparison of vvc, av1 and hevc on 8-bit and 10-bit content," in *Applications of Digital Image Processing XLI*. Intern. Society for Optics and Photonics, 2018, vol. 10752, p. 107520V.

[29] D. Grois, T. Nguyen, and D. Marpe, "Coding efficiency comparison of AV1, VP9, H.265/MPEG-HEVC, and H. 264/MPEG-AVC encoders," in *Picture Coding Symposium (PCS)*. IEEE, 2016, pp. 1–5.

[30] T. Nguyen and D. Marpe, "Future video coding technologies: A performance evaluation of av1, jem, vp9, and hm," in *2018 Picture Coding Symposium (PCS)*. IEEE, 2018, pp. 31–35.

[31] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions," in *Proc. of the 3rd Intern. workshop on video processing and quality metrics*, 2007, vol. 4.

[32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, April 2004.

[33] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Asilomar Conf. Signals Syst. Comput.*, 2003, pp. 1398–1402.

[34] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec 2005.

[35] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. on Broadcasting*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[36] P. V. Vu, C. T. Vu, and D. M. Chandler, "A spatiotemporal most-apparent-distortion model for video quality assessment," in *IEEE ICIP*, Sep. 2011, pp. 2505–2508.

[37] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "The NETFLIX tech blog: Toward a practical perceptual video quality metric," `http://techblog.netflix.com/2016/06/toward-practical-perceptual-video.html`, note = [Online; accessed 2018-08-04],.

[38] Z. Wang, H. Sheikh, and A. Bovik, "No-reference perceptual quality assessment of jpeg compressed images," in *IEEE ICIP*, 2002, pp. 477–480.

[39] H. Sheikh, A. Bovik, and L. Cormack, "No-reference perceptual quality assessment of jpeg compressed images," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, 2002.

[40] S. Gabarda and G. Cristobal, "Blind image quality assessment through anisotropy," *J. Opt. Soc. Am*, vol. 24, no. 12, pp. B42–B51, 2007.

[41] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, 2010.

[42] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 838–842, 2015.

[43] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2013.

[44] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, 2011.

[45] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, 2012.

[46] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind prediction of natural video quality," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1352–1365, 2014.

[47] J. Korhonen, "Two-Level Approach for No-Reference Consumer Video Quality Assessment," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 5923–5938, 2019.

[48] A. Segall, V. Baroncini, J. Boyce, J. Chen, and T. Suzuki, "Joint call for proposals on video compression with capability beyond hevc," October 2017.

[49] A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos, "Using deep neural networks for inverse problems in imaging: beyond analytical methods," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 20–36, 2018.

[50] C. Ledig and et al., "Photo-realistic single image super- resolution using a generative adversarial network," in *IEEE CVPR*, 2017, p. 105–114.

[51] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," April 2001.

[52] Fan Zhang, Felix Mercer Moss, Roland Baddeley, and David R Bull, "BVI-HD: A video quality database for HEVC compressed and texture synthesized content," *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2620–2630, 2018.